

Entrapment: From the Objection from Temptation to the Objection from Moral Alliance

ABSTRACT: Entrapment involves two parties: the agent, who entraps, and the target, who is entrapped. When entrapment occurs, the agent procures the target's act by recommending, requesting or enjoining its performance. When an agent intentionally tempts a target, the agent presents the target with the opportunity to perform an act, with the intention that the target will experience an urge to perform it. Mere virtue testing is weaker: here, the agent intentionally presents the opportunity, but without the intention to tempt. One moral objection to entrapment has it that the entrapment is impermissible because it involves temptation. With the benefit of the aforementioned background material, we scrutinize, and reject, both this objection and a logically weaker one that is recoverable from it. We argue, instead, that when entrapment morally allies the agent, in a manner that we explain, with the target's act then it is thereby impermissible.

KEYWORDS: culpability, entrapment, ethics, procurement, temptation, virtue testing

Introduction

Cases of entrapment involve a party that intends to entrap, whom we call the ‘agent’, and a party that is entrapped, whom we call the ‘target’. In what follows let the terms ‘party’, ‘agent’ and ‘target’ encompass both individuals and groups. The terms ‘agent’ and ‘target’ will also be used in cases of mere temptation and mere virtue testing. Whether a person is tempted primarily concerns their emotions, rather than their actions. When a target is tempted, we take it, the target experiences an urge to commit an act, whilst also being to some degree internally conflicted about that urge (cf. Hughes 2004; 2006a; 2006b). If a person feels tempted and acts in the satisfaction of this urge, then that person has succumbed to temptation. We distinguish between *agential temptation*, in which an agent succeeds in the agent’s designs to tempt a target (even if the target does not succumb), and *situational temptation*, in which a person finds himself or herself in circumstances that (whether by accident or by design) the person interprets as providing an alluring opportunity. A situation may also be said to be tempting when it would tend towards the situational tempting of persons put into it. Whereas the situational temptation of an agent is an occurrent matter, that a situation is tempting in the sense just mentioned is a dispositional fact about it. Except in passing, we are not concerned with situations that are merely situationally tempting, whether as an occurrent or as a dispositional matter. We limit our discussion to those scenarios in which an agent intentionally presents a target with the opportunity to perform an act that the agent recognizes to be legally or morally impermissible.

In their discussion of entrapment, Miller and Blacker (2005: 102–109) make a useful distinction that is pertinent also to the cases of virtue testing and of entrapment. This is between *targeted* action by an agent, aimed at some target in particular, and *random* action, aimed at no-one in particular. For example, if an agent approaches a person whom the agent already intends to perform an act, and attempts to procure that act, then the agent’s act of

attempted procurement is a targeted one. If, on the other hand, the agent arranges a posting to a billboard encouraging people who happen to pass by to perform an act, then the agent's act of attempted procurement is random. Except when precision requires otherwise, and we say so, we intend the terms 'entrapment', 'temptation' and 'virtue testing' to encompass both targeted and random acts. When providing definitions, we deliberately word them so as to admit both targeted and random readings.

Virtue testing, as we define it, occurs when an agent intentionally and knowingly presents a target with an opportunity to perform an act that the agent recognizes as impermissible, so as to discover whether the target is disposed to perform it. On this account, virtue testing is not, *per se*, a form of entrapment. (Contrast here the taxonomies of, for example, Dworkin 1985 and Miller and Blackler 2005: 102–109.) When an agent intentionally and knowingly *tempts* a target, the agent presents this opportunity with, *furthermore*, and among other things (see Section 2), the intention that the target will experience an urge to perform the impermissible act. (On the distinction between virtue testing and tempting, compare Hughes 2004: 56.) Virtue testing is consistent with, but does not entail, agential temptation. For two parties, *A* and *B*, if *A* tests *B*'s virtue then it does not follow that *A* tempts *B*, or even that *A* intends to tempt *B*.

Entrapment into an impermissible act happens when the target performs an act that the agent recognizes to be impermissible, and that the agent has procured. (We argue, in Section 3, that if *A* entraps *B* then this does not entail, but is consistent with, its being the case that *A* tempts *B*.) In scenarios of virtue testing, of agential temptation and of entrapment, a common element is the intentional presentation to the target, by the agent, of the opportunity to perform an act that the agent recognizes as impermissible. That is, the agent puts the target into a situation that the agent intends the target to interpret as providing a certain opportunity.

The verbs 'test', 'tempt' and 'entrap' are verbs of success: if, for example, an agent's intention to entrap is not fulfilled then we do not have entrapment, but attempted entrapment.

While we agree with Day (1993: 176) that ‘tempt’ is a verb of success, we disagree with him, and agree with Hughes (2004: 49) about the nature of the success. Day holds whenever *A* tempts *B* to *V* it follows that *B* actually does *V*. This obliterates the distinction between success in the intention that the target should experience the urge to *V*, i.e., on our account, success in tempting the target, and success in the intention that the target should yield to the temptation. Contrary to Day’s account, there is no contradiction in utterances of the form ‘*A* tempted *B* to *V*, but *B* did not succumb’.

In Section 1, we summarize our account (from REDACTED), of the concept of entrapment. This account is, of itself, neutral about the question of whether entrapment is morally permissible.

In Section 2, we explain an initial objection to the moral permissibility of entrapment that we call ‘the objection from temptation’. We discuss, and reject, the version of this objection that is proposed by Hughes (2004). On his account, the wrong of entrapment resides mainly in the fact that in tempting the target, the agent compromises or undermines the autonomy that is necessary to the target’s properly being held culpable for the entrapped act. After explaining why we regard Hughes’s version of the objection as unsound, we strip the objection down into a more minimalistic, logically weaker and philosophically more plausible version. According to this version of objection, the agent intends that the target should fall to the temptation to perform an act recognized to be impermissible by the agent; it is impermissible to act on the intention that an act that one recognizes as impermissible be performed (whether by oneself or another); thus, entrapment is impermissible.

We argue, by contrast, that not all cases of entrapment are cases of temptation. While there are cases of entrapment in which the agent succeeds in tempting the target, there are also cases in which the agent does not intend to tempt, and cases in which, while the target performs the impermissible act, it would be incorrect to describe the target as having been tempted (as

against merely motivated) to perform it. This significantly limits the scope of the objection from temptation, both in the more complicated form given by Hughes (2004) and in the simpler form arrived at in the course of our critique of his position. We point out that the appeal to temptation is, in the simpler version of the objection, in any case redundant.

Removal of the appeal to temptation leaves us with an even more sparse argument in objection to entrapment, that we call ‘the objection from intention’. It locates the wrong-doing of those acts of entrapment that are of present interest (see Section 1) in the agent’s intention that the target should act in a manner recognized as impermissible by the agent (regardless of whether the target is, along the way, tempted). According to the objection, the impermissibility of the target’s entrapped act spreads to the agent’s act of entrapment.

In Section 3, we shift our critical attack onto the only philosophically interesting or contentious part of the objection from intention. This is an ethical principle that we call ‘the Purist Principle’. It says that if one recognizes that an act is impermissible (that is, one regards it as impermissible within a legal or moral system that one regards as binding, and one is correct about its impermissibility within that system) then it is impermissible for one to intend its performance (whether by oneself or another). By appeal to various counterexamples, we advocate rejection of the Purist Principle.

Section 4 begins by drawing a conceptual distinction between the intentional presentation of the opportunity to perform an act that one recognizes to be impermissible, and the procurement of such an act. We explain how an agent can intentionally present a person with an opportunity to act in an impermissible way but without intending that the person take up the opportunity. We draw a moral distinction between such relatively innocuous, albeit sometimes morally risky, scenarios and scenarios in which, more seriously, the agent does intend that the target should take up the opportunity. Procurement involves the latter sort of intention.

In Section 5, we salvage what we take to be a kernel of real moral insight that lies behind the Purist Principle, and we turn it in to a new moral objection to entrapment. We call this new objection ‘the objection from moral alliance’. According to it, typical cases of entrapment are impermissible because an agent who entraps thereby becomes morally allied, in a manner that we explain, with the target’s impermissible act. We explain a restricted class of acts of entrapment that we take to be immune to this objection, but which do not threaten its applicability outside of that class. We then explain why our stance on entrapment does not commit us to any stance on the admissibility or otherwise of evidence gained from entrapment in court.

In Section 6, we conclude our discussion of the moral permissibility of entrapment.

1. Entrapment

The material in this section develops that in REDACTED. We draw two distinctions, which cut across each other, concerning acts of entrapment. The first concerns the status of the agent and the second concerns the act that the target commits and that the agent has procured.

Our first distinction concerns the agent’s status. *Legal entrapment* occurs when the agent is a law-enforcement officer, acting (in accordance with the law or otherwise) in an official capacity as a law-enforcement officer, or when the agent is acting on behalf of a law-enforcement officer, as the officer’s deputy. (Some jurisdictions may well specify that any act done that is not done in accordance with the law is not an act of a law-enforcement officer acting in their official capacity. We allow, nevertheless, for the possibility of an officer’s abusing their powers while still acting as an officer.) When, on the other hand, the agent is neither a law-enforcement officer acting in that capacity nor the deputy of such an officer, acting as a deputy of the officer, we have *civil* entrapment.

Our second distinction concerns the procured act: we distinguish between acts that are of a criminal type and those that are not. For an example of the latter case, an investigative journalist might entrap a politician into committing a morally compromising act that is not a crime, in order that the journalist can expose the politician for having committed the act. When the act is non-criminal but is morally compromising in some way (whether by being immoral, embarrassing, or socially frowned upon in some way), we are dealing with *moral entrapment* (using the word ‘moral’ in a wide sense). When the act is of a criminal type, we have *criminal entrapment*.

Thus, four types of entrapment can be distinguished: legal criminal entrapment (e.g., the police entrap someone into dealing in illegal drugs), civil criminal entrapment (e.g., a journalist entraps someone into dealing in illegal drugs), civil moral entrapment (e.g., a journalist entraps a politician into making an embarrassing boast) and legal moral entrapment (e.g., the police, in their capacity as law-enforcement agents, entrap a politician into making an embarrassing boast). In what follows, we restrict our concern to cases of entrapment in which the agent intends to entrap the target into performing an action that the agent considers to be, and which is, morally impermissible: i.e., to those cases in which the agent *recognizes* the act that the agent intends the target to perform as morally impermissible. Such cases, we take it, may be drawn from all four of these types of entrapment. Entrapment occurs whenever:

- (i) an agent plans that a particular act be committed;
- (ii) the planned act is of a type that is criminal, immoral, embarrassing, or socially frowned upon (measurable in part by the extent to which the target would probably not like the act to be exposed to colleagues, an employer, friends, family, or the public);
- (iii) the agent procures the act (by solicitation, persuasion, or incitement);

- (iv) the agent intends that the act should, in principle, be traceable to the person performing it (the target) either by being detectable (by a party other than the target) or via testimony (including the target's confession), that is, by evidence that would link the target to the act;
- (v) in procuring the act, the agent intends to be enabled, or intends that a third party should be enabled, to prosecute or to expose the target for having committed the act.

We provide an extended discussion and defence of these conditions, which build on the work of Stitt & James (1984), Gerald Dworkin (1985) and Ho (2011), in REDACTED. The wording of (i) here differs a little from how we phrased it there. This is to make it clear that it is not necessary for the agent to have the *de re* intention that a particular target will perform the act.

We intend condition (v) to include blackmail cases in which the agent intends not that the target will be prosecuted or exposed but that the target will be placed under threat of prosecution or exposure.

We stress that our definition of entrapment makes use of a very specific understanding of procurement. We understand procurement to involve the agent in having an intentional influence, via directly related communicative acts, on the target's will. If the agent merely intentionally presents the agent with the opportunity to perform the act, then this does not amount, on our account, to entrapment.

For reasons explained in REDACTED, our definition of entrapment is neutral about whether the target is culpable for the act and about whether the agent, in entrapping, acts permissibly. In the rest of this article, we are primarily concerned with the question of permissibility.

2. The Objection from Temptation

There are many moral objections to entrapment in the literature: Howard (2016: 25–8) provides a brief survey, with critical evaluation, of eight such objections. (Some of these are relevant only to some of the types of entrapment set out above.) Howard (2016: 26, his *italics*) states the first objection in his list as follows:

Culpability. Entrapped defendants, like persons who have been coerced or drugged, may not be *culpable* for their offences; entrapment morally exculpates offenders, rendering their punishment unjustified.

The objection from culpability, as Howard literally states it, is specific to cases of criminal entrapment. However, it could easily be generalized, so as to encompass all four types of entrapment covered in Section 1, above, as follows:

Entrapped parties, like persons who have been coerced or drugged, may not be *culpable* for the impermissible acts that they have been entrapped into performing; entrapment morally exculpates them, rendering unjustified whatever punishment, opprobrium or ostracism might, due to these acts, be directed towards them.

Howard (2016: 26, note 3) rightly sees Hughes (2004) as an exponent of the objection from culpability (as it is stated by Howard). The objection that we will take as the stimulus for our discussion, which we call ‘the objection from temptation’, is related to, but distinct from, this objection from culpability.

According to the objection from temptation, entrapment is morally impermissible either because it involves agential temptation, or because it is significantly analogous morally speaking, to agential temptation. (The objection is really, as will shortly emerge, a family of objections.) Arguments of the first type are *directly from temptation*. Those of the second are *by analogy with temptation*.

We may distinguish between versions of the objection from temptation. (An act can be wrong for more than one reason, and instances of these two versions of the objection can be, but need not always be, consistent with each other.) A *target-centred* version has it that entrapment is wrong because of the effects that the agent's act of entrapment has on the target's mental states or agential capacities. (Hughes (2004) defends a target-centred version of the objection from temptation. Shortly, we discuss this further.) An *agent-centred* version, which we discuss in more detail in the next section, has it that entrapment is wrong because the entrapping agent has the impermissible intention that the target will succumb to the temptation to perform an impermissible act. It is the focus on the acts or intentions of the agent that makes a version of the objection from temptation agent-centred. The specific agent-centred version of the objection that interests us works with the general principle that if an act is impermissible then it is thereby impermissible that one should intend its performance by another.

The basis of the objection from temptation, then, is that there is, in relevant circumstances at least, something wrong with agential temptation. According to the objection, entrapment, at least of a given sort, involves agential temptation, and is wrong for that reason, or else, if it does not involve agential temptation itself, it is wrong because it shares the feature that makes agential temptation wrong. Hughes (2004) argues that legal criminal entrapment, when it is without 'probable cause', is impermissible for the first of these reasons: that is, because it involves agential temptation. (Among the many other authors who

take it that any act of agential entrapment involves agential temptation, but who do not explicitly offer this as a moral objection to entrapment are Park (1976: 270–273) and Ho (2011: 72–73).)

According to Hughes (2004: 45, cf. 55), this sort of ‘entrapment [...] illegitimately violates’ the target’s moral autonomy. For Hughes (2004: 54–55), an agent who entraps in this way acts wrongly because that agent deliberately subjects the target to temptation. Whenever an agent tempts a target, then, in turn, the target’s ‘will is locked in an internal conflict, and this compromises or sometimes even undermines [the target’s] autonomy’ (Hughes 2004: 53). In cases of legal criminal entrapment without probable cause, it is this sort of manipulation of the target’s will that makes the entrapping agent’s act wrong (Hughes 2004: 56). As Hughes (2006a: 224, *his italics*, cf. 2006b: 355) summarizes his position:

entrapment, like temptation in general, compromises or even undermines the autonomy of those entrapped. Since autonomy is required for reasonable ascriptions of moral and legal responsibility, the *moral* problem with entrapment is that it defeats an essential condition of criminal liability in the process of detecting and preventing crime.

A version of the objection from temptation underlines Hughes’s support for the objection from culpability (in relation to criminal entrapment). The position that Hughes endorses is as follows. Entrapment must involve the agential temptation of the target by the agent. This temptation compromises or undermines the target’s autonomy, in a manner that is incompatible with the target’s culpability for the entrapped act. A generalized version of the objection from temptation can thus function as an underlying basis for the generalized version, set out above, of the objection from culpability. For any entrapment scenario, from

whichever of the four types discussed in Section 1, the view would be, these premises hold. Thus, given the absence of culpability, no punishment, opprobrium or ostracism of the target would be justifiable. So, Hughes (2006a: 228) is correct in his suggestion, which he does not there explain, that his version of the objection from culpability for cases of criminal entrapment readily extends to other cases of entrapment. (Hughes's ultimate position, however, does not debar all acts of entrapment. Hughes (2004: 57) holds that legal criminal entrapment *with* probable cause is permissible but, as far as we can tell, he merely asserts this view without argument. The view seems hard to reconcile with his wider discussions of entrapment and of temptation.)

A generalized argument from temptation that is recoverable from Hughes's work goes as follows.

1. Whenever an agent, *A*, entraps a target, *B*, *A* intentionally tempts *B* to perform the entrapped act, for which *A* intends that *B* will be held culpable. Premise
2. When *B* is so tempted, *B*'s autonomy is either compromised or undermined. Premise
3. One is culpable for an act that one has performed only if, in performing the act, one's autonomy is not compromised or undermined in this way. Premise
4. If a person is not culpable for an act, then it is wrong to hold that person culpable for it. Premise
5. If an act is wrong, then it is wrong to act with the intention that it be performed (whether by oneself or by another). Premise

- | | | |
|----|--|-----------------|
| 6. | Whenever an agent, <i>A</i> , entraps a target, <i>B</i> , then <i>B</i> is not culpable for the entrapped act. | From 1, 2,
3 |
| 7. | Whenever an agent, <i>A</i> , entraps a target, <i>B</i> , then it is wrong to hold <i>B</i> culpable for the entrapped act. | From 4, 6 |
| 8. | Whenever an agent, <i>A</i> , entraps a target, <i>B</i> , then <i>A</i> acts wrongly. | From 5, 7 |

Before we return to the specifics of Hughes's own position, let us evaluate some aspects of this more general argument, which he does not himself give. Performing this task serves two roles: it will figure in our evaluation of Hughes's own argumentation and it will facilitate the transition into the next phase of our own argument.

The argument is valid and we grant Premise 1. Premise 4 is ambiguous. It is a known feature of systems of criminal justice that the innocent are sometimes mistakenly convicted of crimes. When a court convicts someone who is, but who is not known to be, innocent, but it does so unknowingly, then the court does not necessarily act in an impermissible way.

Provided that due process has been followed, and that the system of criminal justice is not corrupt, the court acts permissibly. If the wrong verdict was arrived at, then this can be an epistemic mistake, and it is something that would not happen in a morally ideal world, but the court does not necessarily act impermissibly. Premise 4 is acceptable only if 'wrong' in it is interpreted to mean 'incorrect' or 'mistaken', rather than 'morally wrong' or 'impermissible'.

To get rid of the ambiguity in 4, we can recast the argument as follows, with attendant changes to the wording of 5, 7 and 8:

- | | | |
|----|--|-----------------|
| 1. | Whenever an agent, <i>A</i> , entraps a target, <i>B</i> , <i>A</i> intentionally tempts <i>B</i> to perform the entrapped act, for which <i>A</i> intends that <i>B</i> will be held culpable. | Premise |
| 2. | When <i>B</i> is so tempted, <i>B</i> 's autonomy is either compromised or undermined. | Premise |
| 3. | One is culpable for an act that one has performed only if, in performing the act, one's autonomy is not compromised or undermined in this way. | Premise |
| 4. | It is impermissible to hold a person culpable for an act whom one knows not to be culpable for it. | Premise |
| 5. | If an act is impermissible, then it is impermissible to act with the intention that it be performed (whether by oneself or by another). | Premise |
| 6. | Whenever an agent, <i>A</i> , entraps a target, <i>B</i> , then <i>B</i> is not culpable for the entrapped act. | From 1, 2,
3 |
| 7. | Whenever an agent, <i>A</i> , entraps a target, <i>B</i> , then it is impermissible for any party <i>C</i> , who knows that <i>B</i> was entrapped, to hold <i>B</i> culpable for the entrapped act. | From 4, 6 |
| 8. | Whenever an agent, <i>A</i> , entraps a target, <i>B</i> , then <i>A</i> acts impermissibly. | From 5, 7 |

Premise 4 now seems unobjectionable. The problem is with its interaction with the other premises, especially the particularly controversial Premises 2 and 3. Supposed that *A* has entrapped *B*. *C*, who knows this, holds *B* culpable for the entrapped act. *C* might also accept Premise 4. *C* will not necessarily thereby be in breach of Premise 4, for the following three reasons. First, *C* might, contrary to Premise 1, not think that all entrapment scenarios involve

temptation, and may believe that *A* entrapped *B* without having tempted *B*. Shortly, we will argue that not all entrapment scenarios involve temptation. Secondly, *C* might grant that whenever entrapment happens so does temptation, but hold, contrary to Premise 2, that the temptation involved need not seriously compromise or undermine the target's autonomy. We think that this is a correct response to Premise 2, because, on our account, when *A* entraps *B* then *A* inclines *B*'s will, through solicitation, incitement or persuasion, towards performance of the entrapped act. Such inclination of the will is not, unlike, for example, the inclination of *B*'s will through administration of a potion, a case of compulsion. Even if there are temptations that are irresistible, and thus which compel, to tempt is not, *per se*, to compel. Since compromise to autonomy is, like culpability, presumably a matter of degree, Premise 3 is too strong. Even if entrapment involves temptation, and if it thereby involves compromise to *B*'s autonomy, compromise is not the same as nullification. To the extent that *B*'s autonomy was not compromised, and in the absence of any other argument that exonerates *B*, *B* remains culpable. (Cf. Fletcher (1978: 542), Altman and Lee (1983: 59), Dillof (2004: 895).)

These considerations show, we think, that neither the generalized version of the target-centred objection from temptation nor the more restricted version of the objection actually offered by Hughes himself is sound.

We now shift our focus, therefore, to an agent-centred version of the objection that is far simpler. It is encapsulated in the following short argument:

1. If an act is impermissible, then it is impermissible to act with the intention that it be performed (whether by oneself or by another). Premise
2. In entrapment, the agent intends that the target should fall to the temptation to perform an impermissible act. Premise

3. Entrapment is impermissible.

From 1, 2

This agent-centred argument is recoverable from the previous, target-centred, one. Premise 1 here is the same as Premise 5 of the previous argument, while Premise 2 is derivable from Premise 1 of the previous argument. Since the premise set of the agent-centred argument is a proper subset of that of the target-centred argument, the agent-centred argument is logically weaker. Given that both arguments are valid, any refutation of the agent-centred argument will spread to the target-centred argument. Accordingly, even those who, unlike us, accept Hughes's views on the relationship between temptation and culpability will need to reject his version of the objection from temptation if, as we do, they find fault with this agent-centred argument.

In the definition of entrapment that we set out in Section 1, the notion of temptation played no role. On our account, in entrapment scenarios the agent's crucial intention is that the target perform the act that the agent is trying to procure. What goes on the target's mind, other than inclination of the target's will towards performance of that act, formation of the intention to perform it, and execution of that intention, need not be of concern to the agent. (While we take procurement to be a specific form of agential inducement, Miller and Blackler 2005: 102–105, are right, we think, to talk about *inducement* into the entrapped act, rather than the *temptation* to perform it, which they do not even mention. Miller, Blackler and Alexandra (2006: 264) depict inducement as something that a target may 'resist', but they do not explicitly link it to temptation.) Indeed, an agent who agreed with Hughes's views on the relationship between temptation and culpability might just want to procure the act so that the target could properly be held culpable for it, and might intend only that the target perform it, with no ambivalence or mental anguish along the way, rather than perform it as a result of

having been tempted and succumbed. Such an agent wishes the target to be *motivated* to perform the act but is at best indifferent about whether the target is *tempted* to perform it. (It is easy to see that being motivated to *V* differs from being tempted to *V*. We are motivated, for example, to floss our teeth but we are not ordinarily tempted to do it.) If this is correct, then no version of the direct objection from temptation can succeed: even if entrapment commonly involves temptation, it does not always do so. For this reason, among others, we agree with Redmayne (2012: 161), that ‘it is not being subjected to temptation that is really objectionable in entrapment’.

If the following generalization were true, then every case in which a criminal act is performed as a result of entrapment would be a case of yielding to temptation: ‘All crime is a yielding to temptation, the temptation to obtain whatever gains, pecuniary or nonpecuniary, the crime offers’ (*United States v. Manzella*, 791 F.2d 1263, 1269, 7th Cir. 1986). A little thought shows that the generalization is fundamentally flawed, and no threat to the above argument. A hardened criminal who commits several crimes a day may do it as second nature, without experiencing any inner conflict: thus, the criminal acts in question would *not* be cases of yielding to temptation.

3. The Purist Principle and the Objection from Intention

If we strip away from the agent-centred version of the objection from temptation the appeal to temptation, we are left with the following valid argument, which we call ‘the objection from intention’.

1. If an act is impermissible, then it is impermissible to act with the intention that it be performed (whether by oneself or by another). Premise

2. In the cases of entrapment that are of interest to us (see Section 1), Premise the agent intends that the target should perform an act that is impermissible, and so regarded by the agent.

3. Entrapment is, in such cases, impermissible. From 1, 2

Premise 2 of this argument is logically weaker than that of the previous argument, and, in any case, it true as a matter of stipulation. It is Premise 1 that is of philosophical interest.

According to the objection, entrapment is impermissible for the same reason that it is impermissible for an agent intentionally to tempt a target to perform an act that the agent recognizes as morally impermissible. Like temptation, entrapment breaches the plausible-sounding principle, which we have called the ‘Purist Principle’, and which Premise 1 expresses, that if a certain action is morally impermissible then it is also morally impermissible to intend that it be performed. This principle does not require that the agent’s motivating thought be that the target should perform an impermissible action. Rather, it suffices that the agent recognizes that the action is impermissible. Since temptation and entrapment alike involve intending the performance, by the target, of an impermissible action, then, the thought is, they are both impermissible.

To see the initial appeal of the Purist Principle, take the following example. The Mafia boss wants to keep his or her own hands ‘clean’, and so manipulates the easily led tyro into doing the boss’s dirty business and killing a rival that is threatening the boss’s interests. This manipulation need not take the form of full-blown causation; it could be that the boss is a shrewd judge of character, and can predict how the tyro will react in different situations. Nor need these predictions be infallible; if the mafia boss intends that the tyro kill the rival while recognizing that this intention may not be fulfilled, this seems enough to establish that the Mafia boss is morally at fault for having such intentions.

One response to this example is that the Mafia boss need not intend that the tyro have an impermissible intention, or even intend that the tyro perform an impermissible action. Suppose, for example, that the boss can manipulate the tyro into thinking that the tyro must kill the rival for reasons of self-defence. In that case the boss need not intend that the tyro have an impermissible intention (since the intention to use lethal force in self-defence need not be impermissible) and need not intend that the tyro perform an impermissible action (since a killing committed in what the killer thinks is self-defence need not be impermissible). Nevertheless, the boss intends that the bad state of affairs of the death of the rival obtain, and this intention seems impermissible.

Although this line of argument in favour of the Purist Principle might sound plausible at first, we hold that it is mistaken. We will now set about refuting the Purist Principle, by counter-example.

In the well-known trolley problem, first introduced by Foot (1967), the bystander at the switch has the opportunity to switch the points to direct a runaway trolley (a tram-like street-car) away from its main line, on which there are five innocent people, to its branch line, on which there is just one innocent person. The common view, which we accept, is that it is morally permissible to redirect the trolley. Here the redirection is spatial, from main line to branch line. What, however, if the redirection were temporal? What if the trolley were set to roll in the future at a time at which there would be five innocents in its path, but it could be set to roll later or earlier, at a time at which there would be only one innocent in its path? It seems unproblematic to set it to roll later. We also suggest that it is unproblematic to set it to roll earlier: the uncomfortableness one feels with this suggestion is merely due to the assumption that the later it is in the future the more likely one will be able to stop it altogether or save all the people.

Now, the standard version of the trolley problem does not feature an evil human agent, but an out-of-control trolley. Nevertheless, we think that the same principle would obtain in what we might call ‘the evil driver version’. In this version there is an evil driver in control of the trolley. The evil driver can stop the trolley at any time, but is driving it down the tracks towards the five intending to kill them. The driver is ignorant of the presence of the one on the branch line. Now, if the bystander at the switch is merely switching the points, then the bystander is not causing any change in the conscious decision-making process of the driver. But it seems to us that the same considerations would apply if the bystander were causing changes in the driver’s conscious decision-making process. In other words, it seems to us permissible to attempt to change the driver’s intentions from driving down the main line towards the five to driving down the branch line instead, even though that course of action will result in the death of the one. Now, this does not so far imply that it is permissible to change the evil driver’s intentions towards *aiming at* the one, which intention would be morally impermissible. (It might be responded that the intention to aim at the one is not impermissible, because it is the lesser evil in this context. But note first that in the standard trolley problem the bystander does not aim the trolley at the one, but merely aims it away from the five, and, secondly, that in our version the evil driver does not have the intention to *aim at the one rather than the five*, but the categorical intention to *aim at the one*.) Suppose, however, that the five on the main line are the family of the one, and that the one on the branch line, of whose presence the evil driver is unaware, is the driver’s hated adversary. Is it permissible for the one to alert the driver to the one’s own presence on the branch line in order to get the driver to turn away from the five? We hold that it is permissible for the one to do so, even though the one’s intention is to form an impermissible, murderous intention in the mind of the evil driver. (We can suppose that the one knows that nothing less than that murderous intention will suffice to turn the driver away from attempting to murder the five.)

So, here the one intends to produce in the evil driver an intention that, we may suppose, the one recognizes that it is impermissible to have, namely the intention to murder the one. Yet this seems, in the circumstance, permissible, even though the Purist Principle debar it. What is more, it seems *obviously* permissible if the one knows that the driver will be unable to fulfil the murderous intention. Perhaps, for example, the one knows that the one will be released from the branch line before the trolley can reach the one. The Purist Principle, by contrast, says that it is impermissible knowingly to intend that someone form an immoral intention, even if there is no possibility of the intention's being fulfilled. (Since forming an intention is itself a mental action, it is covered by the Purist Principle.)

Now, this was, in fact, a case of *spatial* redirection, but we mentioned above that a case of *temporal* redirection would be more similar to entrapment. We shall now consider such a case. Rather than considering another variant on the Trolley Problem, let us consider the real-life case of the Q-boat, suggested to us by REDACTED. Take the real-life example of the British use of Q-boats during the First World War. A Q-boat was disguised as an unarmed boat such as a fishing trawler, sometimes used with the intention of luring a German U-boat into performing a surface attack (Wikipedia, 2019). The U-boat would then be disposed of by the Q-boat's hidden guns. If the U-boat attack were recognized by the British forces as impermissible, perhaps because it was against the moral principles of a just war to attack (what was believed to be) an unarmed fishing trawler, then the Purist Principle would rule impermissible the British use of the Q-boat: it was intended to lead the U-boat commander into a morally impermissible attack. This result, however, seems to us to show that the Purist Principle is incorrect. A possible response, again suggested to us by REDACTED, on behalf of the Purist Principle is as follows. The U-boat commander (like the evil driver of the trolley) is already engaged in a morally illicit enterprise, that of finding unarmed civilian boats to destroy. Therefore, the use of the Q-boat merely bringing to the

surface an already existing, but previously hidden, impermissible action. Since the use of the Q-boat's does not signal the intention, on the part of British forces, that the U-boat commander should perform any new impermissible action, the Purist Principle does not apply. Against this, however, the U-boat is not already engaged in an attack when it first encounters the Q-boat. Thus, it seems to us that the use of Q-boats does signal the intention that U-boats should start new attacks. After all, the U-boat might indeed not be looking for a target at all when it stumbles across the Q-boat. It should be noted that there need be no spatial redirection involved here: it may be that if the Q-boat does not draw it up the U-boat will stay in more or less the same spot until an unarmed passenger ship is sighted. There is, however, temporal redirection: the British forces know that each U-boat commander will, at some point, if presented with the opportunity, launch an immoral attack. The use of the Q-boat serves to prompt the U-boat commander to begin his immoral attack now, rather than at a subsequent time when the Q-boat is not there and, for example, a real unarmed passenger ship may be in the U-boat's sights. We conclude that the case is similar to the Temporal Trolley Problem. To use a Q-boat for the purpose for which it is intended is to redirect an evil from the future, where it could be disastrous, to the present, where it can be neutralized.

Another, more homely, example of temporal redirection can be seen here: suppose that a child's grandparents have an unhealthy adoration for him. The parents wish the grandparents to realize that the child is not perfect, so they put a tempting situation in front of the child (e.g., they leave a marshmallow out between meals when the child knows that marshmallows are forbidden between meals), in order that the child might do wrong and the grandparents might be less uncritical in their adulation. The Purist Principle once again rules out this action on the parents' part, yet it seems to us permissible for the parents to intend that their child do wrong, which shows that it is not a general principle that it is wrong to intend that someone do wrong. (Note that the relevant contrast here is not between oneself and

another. There may be circumstances in which one intends that one will oneself do wrong at a later stage, e.g. one might intend to have a quarrel with one's beloved and to make up afterwards, thinking 'The sequence of having a quarrel and then making up afterwards is better than just not having a quarrel at all, so I intend that I bring about a quarrel down the line in a week or so in order that there be another satisfying reconciliation'.) The temporal redirection should be clear here: the parents know, or reasonably suspect, that their child will do wrong at some point in the future, but want this event to happen, if it does happen, in front of the grandparents. To guarantee this, however, the parents need to provoke the child to do wrong at a certain point, rather than relying on chance.

There is, however, a possible riposte: is it just that the parents *hope*, but don't *intend*, that the child yield to temptation? Suppose, however, that the parents actually increase the temptation level (by, say, providing two marshmallows). In this case it looks as though they really are intending that the child accede to the temptation and do the forbidden thing—else why increase the temptation?

Suppose, for another case, that A is a prominent religious official trying to choose a new pastoral assistant, and that A wishes above all to appoint the morally best candidate out of the many that apply. A has a shortlist and adapts a shortlisting strategy from that used by the special forces and televised reality competitions: A gets the candidates to do more and more challenging tasks (street preaching to rowdy unbelievers, counselling someone tragically bereaved, preparing a sermon at drastically short notice, chairing a meeting of the religious council) until some of them betray their sinful selves by swearing, blaspheming, assaulting A etc. These candidates are then eliminated (from the short list). The problem is that the candidates are better than A expected, and it costs money to keep the short-listing going, so A has to keep increasing the difficulty of the challenges in order that the less holy candidates will do something impermissible and get eliminated. The Purist Principle would

rule such a course of action impermissible, but A is merely trying to redirect into a controlled environment in the present a morally impermissible reaction that would otherwise have happened in the future in an uncontrolled environment, so as to eliminate candidates that would have reacted later in a morally impermissible way in that uncontrolled environment and thereby removing the possibility of that undesirable occurrence. Part of our intuition is that, as in cases of (standard) entrapment, one is not, in a sense, seeking to add to the sum total of impermissible actions in the world. Rather, one is just trying to get the evil lurking in the breast to be expressed. That seems to make a considerable moral difference—if the agent thinks that the target is not antecedently inclined to do something impermissible then it seems that a greater level of justification is required to justify the intention that the target do something impermissible. Nevertheless, we insist that it is permissible for A to intend that the candidates, or a particular candidate suspected of being a ‘weak link’, do something impermissible.

We trust that these examples will suffice to cast doubt on the Purist Principle, and the objection on which it is built. We present our positive theory later.

4. Procurement and the Intentional Presentation of an Opportunity

What is it to procure? One necessary condition of procurement is causation; if one procures an offence one causes it, albeit indirectly, to occur. Not all cases of causation are cases of procurement: if *A* brainwashes *B* into perform an action then *A* has, we say, not procured that action from *B*. By contrast, if *A* threatens *B* by saying that *A* will shoot *B*, or expose *B*’s darkest secrets, unless *A* performs a certain action, then *B* has procured that action. Similarly, if *A* bribes *B* by saying that *A* shall pay or look out for *B* in the future, this counts as procurement. Procurement always goes by appeal, in a rational way, to the will, even if the target acts

irrationally in performing the procured action. Procurement need not involve an inducement of pleasure to be had or pain to be avoided. One can procure an action merely by recommending it or requesting it. The phrase ‘recommending it, or requesting it’ must be construed so as to include heavy hints and the use of implicature to recommend (as in ‘It would be good if our competitor could be made to disappear’). Recommendations and requests can also be communicated via such methods as gestures, drawings, mime, and sign-language.

We take it that procurement involves *intentional* causation. If one accidentally or unintentionally causes someone to perform a wrongful act, then that is not a case of procurement. (It does not follow that in procuring one always intends that one *procure*: many people engaged in procurement are perhaps quite unaware that there is such a thing as procurement, and are fully concentrated on the causation.)

Let us apply this to a couple of examples from earlier. If the agent leaves a wallet lying on the pavement in order to trap the target, then the agent has not procured the act unless the agent in some way recommends the action to, or requests the action of, the target. It is clear that if an agent actually asks a suspected drug-dealer to sell drugs then that is a case of procurement. By contrast, if the agent just hangs around hoping that the drug-dealer will approach and offer the drugs for sale then that is not procurement, even if the agent uses a disguise to resemble a potential client.

This account of procurement is somewhat narrower than the standard one used by the courts of England and Wales. In *Attorney General's Reference (No. 1 of 1975)* [1975] QB 773, Lord Widgery defined procurement as follows: ‘To procure means to produce by endeavour. You procure a thing by setting out to see that it happens and taking the appropriate steps to produce that happening’ (*Attorney General's Ref.* 779F). He allows to qualify as procurement of drink driving the surreptitious lacing of a drink without the driver’s knowledge. On our account this would not qualify as an example of procurement.

Even supposing, against our own account, that entrapment always involves temptation, there is a difference between being entrapped into the performance of an act and performing it as a result of mere temptation. In entrapment the agent seeks to bring about the performance of the act by procurement, that is, by means of certain directly related communicative acts. For example, the agent might ask the target ‘Do you have any drugs for sale?’. Or the agent might use the imperative form ‘Please sell me some drugs!’ or might make a statement with an obvious implication such as ‘I’d like to buy some drugs’. The meaning of ‘directly related’ should be clear in view of the contrast that we now draw.

If, on the other hand, the agent seeks to bring about the performance of the act by means of other, non-communicative, wiles, or by indirectly related communicative acts, then we do not have a case of entrapment, even if the agent succeeds in tempting the target into the performance of the act. For example, suppose that the agent is posing as a senior citizen with the intention of luring out a mugger. The agent might spy a target and then try to tempt the target to mug by non-communicative schemes, such as adopting a dodderly walk, ‘accidentally’ dropping some cash, adopting the air of one lost etc. Or the agent might even engage in various communicative acts, such as saying to the target ‘I’m lost and am looking for the bank; can you help?’. Here the agent neither requests nor orders the target to mug; rather, the agent’s communicative acts are intended to create in the target’s mind certain beliefs that are meant to elicit an intention, on the target’s part, to mug the agent (such as the belief that the agent is vulnerable and has money on their person).

In a case of *mere* virtue testing, the agent presents the target with the opportunity for wrong-doing, but without the intention of bringing it about that the target commit, or even be tempted to commit, the wrongful act. For example, suppose that an agent disagrees with a colleague over whether another colleague, Brian, is a thief. The agent might say to the colleague ‘I can prove to you that Brian is not a thief; let’s leave some money lying out as a test, and

you'll see that he won't take it'. Here the agent definitely does not intend that the target take the money; in fact, the agent intends that the target *not* take the money. It might even be the case that the agent is so confident of the target's honesty that the agent does not believe, and does not intend, that the target will experience the temptation to take the money. Nevertheless, the agent intends to provide the target with the opportunity for wrong-doing. We will now argue that it is not always wrong to do this.

Whenever we trust someone to do something, knowing that they might breach our trust, we intentionally present them with an opportunity to do wrong. Doing so is, in itself, innocuous. To trust a trustworthy person is not to tempt them, even though it may be a test for them. In fact, it is sometimes permissible to present someone with the opportunity to do wrong even if one does not think them trustworthy. For example, one might hope against hope that the person passes the test even if one does not think that they will; one might think that everyone deserves an opportunity to prove their moral goodness. So, to revert to an earlier example, it might be that one agent leaves out some money in the sight of a target suspected of theft because the agent thinks the target guilty and wants to demonstrate that fact, while a second agent might leave out some money in the sight of the target because he or she thinks the target innocent and wants to demonstrate *that* fact. In this example, while it seems permissible for either agent to test the target in the way described, it seems *clearly* permissible for the first agent to test the target in the way described even if they think it likely that, regrettably, the target will fail the test.

We do not mean to suggest that it is always permissible for an agent to test a target if the agent intends that the target pass the test. It depends on what the risks and benefits are. If the agent knows that the test is relatively unimportant, but that the target's failure to pass the test—or even the stress of the target's being subjected to the test itself—will cause great harm and upset to the target and many others, then it will usually be impermissible to subject the

target to the test. We do not believe that just any old motive for testing the target will suffice, even if, as it happens, there is a lot at stake. For example, if the agent tests the target just for a laugh, or to see the target squirm under pressure, that would not justify the redirection, even if the world would be a lot better as a result.

In Section 3 we rejected the Purist Principle, but we did not give a positive counter-proposal. We have just stated, what was also implicit in the counter-examples to the Purist Principle, that it is sometimes permissible for an agent to test the target even if the agent intends that the target fail the test. When, then, is it permissible for an agent to intend that the target, by performing a legally or morally impermissible action, fail the test? Typically, this is permissible only if the agent believes that the target is disposed to perform the impermissible action, or something just as bad, in the future, and the agent intends that, as a result of the target's performing the impermissible action now, the impermissible action (or the equally bad action) will not be performed later. This is what is encapsulated in the metaphor of temporal redirection that we deployed in Section 3.

We mentioned two conditions above: standardly, it will be morally permissible for the agent to tempt the target to perform an impermissible action only when it is both the case that the agent intends that the commission of the same or an equally bad action later will be prevented, and the case that the agent believes the target to be antecedently disposed to commit the action (or something just as bad) anyway. These conditions are not absolutely indispensable, however. For example, suppose that the Chief of Police wants to leave money left lying around in order to tempt those police officers likely to steal money. Now, this action is, it seems to us, permissible if the Chief thinks that the officers are likely to steal in the future unless caught now. But suppose that the Chief is not acting from this motive, i.e. is not leaving the money lying around in order to prevent stealing, but solely because proving that the members of his team are totally honest now will attract a good deal of much-needed funding

for the positive aspects of policing, community outreach etc. Here the motivation is not the prevention of an impermissible action, but the creation of a better society. Nevertheless, it seems to us that the intermediate goal of showing that the police force is as honest as it should be is sufficient to justify the Police Chief's actions. Indeed, we think that the Police Chief's action could be justified by the motivation to show that the police force is as honest as it should be, even without the ultimate goal of making the world a better place by gaining the extra funding.

Secondly, we can also imagine circumstances in which it would be permissible for an agent to tempt a target even when the agent does not believe that the target is antecedently disposed to commit the offence. This is because it can be a moral duty to guard not just against expected threats, but even against *unexpected* threats if they are severe enough. (We say that it can be a moral duty, and we assert that in fact it often is a moral duty on law-enforcement agents.) One often has to consider the expected disutility of various outcomes, where disutility is given by a function from the *subjective probability* of the outcome and the *severity* of the outcome. So, it would be permissible to entice the target even if the subjective probability of the impermissible action were only 1%, as long as the threat were severe enough. If the subjective probability were 0%, however, we do not think that it would be permissible for the agent to take action on the basis of a practical impossibility.

Note, in conclusion, that the two standard conditions we have described are independent of one another: it is perfectly possible for the target to be disposed to perform the act if tempted, without its being likely that the target, or anyone else, will perform the act if not tempted. Perhaps, for example, the act is a one-off that it will be impossible for anyone to perform at a later point, and that the target is unlikely to perform even now without being tempted to do so.

Suppose that *A* entraps *B* into the performance of an act that *A* recognizes to be impermissible. *C*, on the other hand, intends to intentionally to present *D* with the opportunity

to perform an act of the same type, recognized as impermissible by *C*, but without intending that *D* should take up the opportunity. Other things being equal, *A*'s act is the more objectionable. This is because in intentionally procuring *B*'s act, *A* becomes allied, in an impermissible way, with that act. This is similar to the common suggestion that lying is worse than mere deception, even if both involve the intention to produce in the other party a false belief: in lying, one allies oneself, by means of one's verbal act, more thoroughly with the false content than one does in mere deception (Carson 2018: 158–159).

An argument for this suggestion can be found in consideration of the reaction of the target on realizing that they have been entrapped. The target may reasonably say 'but you asked me to do it!'. It seems to us that this protest carries force. To see the contrast, suppose that the target has instead been *merely tempted* into performing the illegal or immoral act, i.e. there has been no recommendation of the illegal or immoral act. Here the protest 'but you *intended* me to do it' does not carry the same force, even if the target realized at the time that the agent intended that they perform the action.

How far may one go towards entrapment? For example, is asking the question 'Do you do assassinations?' already going too far? On the one hand, it seems that there is no moral difference between saying 'Do you do assassinations?' and 'Please do an assassination for me', since their perlocutionary effect is the same. On the other hand, a plain inquiry need not betoken a request, and could, in theory, afford the target opportunity to confess and repent rather than to act impermissibly. While it is true that one and the same form of words could be used to quite different effect in different contexts, it seems clear to us that in the context of entrapment a question such as 'Do you do assassinations?' is meant as preliminary to a request for an assassination, and so would be impermissible in that context.

Let us turn our attention to questions of *consent*. Let us consider first the case in which the target consents to being *tempted*. It may be that some employers state as a condition of

employment that the employee may be subjected at some point in the future to virtue testing (somewhat similarly to how beginning athletes are told that they may well be subject in the future to drug testing) designed to flush out the suspected bad apples, and state that if the employee does not consent to this regime then they should not sign the contract.

But can consent justify *entrapment* as well as *temptation*? What if it were a condition of taking a job as a law-enforcement officer that one consent to ‘entrapment testing’, just as athletes have to consent to drug testing to compete at the highest levels? Miller, Blackler and Alexandra (2006: 141) note that policing is an endeavour in which corruption is an ever-present and serious danger, and that ‘testing’ is among the methods of policing used to detect and to deter it. Moreover, they note, ‘detection and deterrence that might not be acceptable in some other professions’ can be justified by the ‘tendency to corruption’ that is, they think, inherent to policing. They add (2006: 142) that the testing may involve targeted or random entrapment. See also their ‘conditions for targeted entrapment’ (ibid.). Given these considerations, there appears to be a strong case for attaching such a condition to employment as a law-enforcement officer.

Would the fact that the target gives a general form of consent make it permissible for the agent to engage in entrapment? While we do not want to hold that one can make just anything permissible by consenting to it, we do think that in the case of entrapment it is possible and permissible for an individual voluntarily to ‘contract away’ their right not to be subject to entrapment, and we can easily imagine situations in which it would be permissible for a law-enforcement agency to make it a condition of employment that employees ‘contract away’ their right not to be subject to entrapment.

5. The Objection from Moral Alliance

We hold that non-consensual entrapment to perform an action recognized as impermissible by the agent is really morally objectional neither because it involves agential temptation nor because of breach of the Purist Principle, but because it allies the agent, via the agent's procurement of it, with an act that the agent recognizes as impermissible. The procurement of the act involves the agent, on our account, in commending, requesting or enjoining the performance of the target's impermissible act. This is structurally similar to the absolutist view against lying. The name 'absolutism' in this context is usually reserved for the view that lying is always impermissible, but there is certainly conceptual space for the view that lying is permissible when someone has consented to being lied to (again, an employee might voluntarily agree to training in how to spot liars, involving practice with some interlocutors lying and some telling the truth), but otherwise impermissible.

On this view, lying is wrong because when one tells a lie one allies oneself with the falsehood that one asserts. In a somewhat similar way, when one entraps one allies oneself with the impermissible action that one procures. Just as the absolutist holds that lying is impermissible, so we hold that (non-consensual) entrapment is impermissible. (We do not think that the proposition that non-consensual entrapment is impermissible implies the proposition that lying is impermissible, or the other way round: it is merely that the structural similarities seem interesting.)

This applies even if the agent does not intend that the target give in to the temptation, and, indeed, even if the agent intends that the target should *resist* the temptation. One might think that it is the agent's intention concerning the target that makes the difference, that it is permissible to counsel or recommend an impermissible action if one intends that the target decline the invitation. (This seems structurally similar to the view that it is permissible intentionally to assert something one believes to be false, if one does not intend that the hearer believe the assertion. Kant (1997: 28) seems to hold this view; for discussion, see

Mahon (2009: 207). We think, on the contrary, that if the agent intends to request or enjoin the impermissible action (without the target's consent) then that suffices to make the agent's action impermissible. (It is possible unintentionally to commend or enjoin something. One can imagine an agent's saying to the target 'do what the superior asked you to do' unaware that the superior asked the target to do something impermissible. There are also, of course, problems with limited linguistic ability where a non-native speaker might ask the target to do something impermissible, unaware that that was the import of the spoken words.) For example, suppose that a principled junior police officer, opposed to entrapment, is instructed by a superior to clear up the problem of drug pushers in a certain area by going undercover and requesting drugs from suspected pushers. Is it permissible for the principled junior officer to obey the letter of the superior's command by going through the motions of entrapment (i.e., saying 'may I buy some drugs?' etc.) with targets that they think are *unlikely* to yield to the temptation? It might be thought that this would be the best of all possible worlds: the principled junior officer is happy that nothing impermissible has occurred, and keeps their job, and the senior officer, in blissful ignorance of the full picture, is happy that drugs have been requested undercover, albeit unsuccessfully. But we demur. We think that even in this case attempting to procure the impermissible action (even if not technically entrapment) is impermissible, since the principled junior officer is still—albeit reluctantly—commending or requesting something, without the target's consent, that they think should not be commended or requested. (Imagine, however, that the senior officer gives this order: 'go and ask "will you sell me some drugs?" of people downtown'. It would be possible for the principled junior officer to avoid entrapment and to follow the *letter* of the senior officer's order by saying the words 'will you sell me some drugs?' in a sarcastic way that made it clear that drugs were not really being requested. In this situation no entrapment would have occurred, even if somebody failed to notice the sarcastic tone, produced some drugs, and was arrested: the

target here would no doubt *feel* entrapped, but would no more *be* entrapped than if they had misheard ‘have you any dogs?’ for ‘have you any drugs?’.)

At this point it might be retorted: why is it not also wrong for the agent to intend that something be performed that they think should not be performed? (This seems to us structurally similar to the complaint that deception is as bad as lying since in deception the deceiver intends that the target believe something that the deceiver thinks should not be believed.) This would be to take us back to the Purist Principle, however, which we rejected earlier. Our contention is that it is the verbal commendation or request, without the target’s consent, that the agent believes to be impermissible that makes entrapment impermissible.

It might also be objected: does not the criminal lose the right not to be entrapped, just as, so the objection might continue, the liar has lost the right not to be lied to? We disagree. We think that there are many rights that one does not lose just through being a criminal. The right not to be tortured is a fairly uncontroversial example. The right not to be raped would be another. It seems to us that the right not to be entrapped is a third, albeit we concede that this is less obvious than in the other examples.

What follows for the entrapped target from our position that entrapment is impermissible? In the United States entrapment is a defence to any charge, so the target would not be convicted in court (Marcus 2016: REF). In England and Wales, and several other common-law jurisdictions (Wells 2011: para 5.22), the judge has discretion to stay proceedings (Wells 2011: para 5.1), and also has discretion to exclude evidence (Wells 2011: para 5.12; Police and Criminal Evidence Act 1984, s. 78).

Our position does not entail that evidence gained impermissibly may be excluded (still less that it must be excluded), that the trial may or should be halted, or that the defendant may or should be found not guilty. We can certainly imagine circumstances, such as in cases of

immanent catastrophe, in which it would be morally permissible, and perhaps even obligatory, for law-enforcement officers to make use of impermissibly gained evidence.

6. Conclusion

We have distinguished in turn between mere testing and temptation, and between mere temptation and entrapment; while we hold that all temptation is testing (in the sense here at issue), we deny that all testing is temptation, we deny that all temptation is entrapment and we deny that all entrapment involves temptation. We have considered at length an initially attractive position, the Purist Principle, that it is impermissible to intend that someone perform an act that one recognizes as impermissible, and we have rejected this principle. We have argued that while it is in many circumstances permissible to test a target, and in some circumstances permissible to tempt a target, it is in no circumstances permissible to entrap a target into doing something impermissible without the target's previously having consented, in a general way, to be subjected to attempts at entrapment.

Acknowledgements

REDACTED.

Cases

United States v. Manzella, 791 F.2d 1263 (7th Cir. 1986).

Table of Laws

Police and Criminal Evidence Act 1984

References

- Altman, Andrew and Steven Lee (1983) 'Legal Entrapment', *Philosophy and Public Affairs* 12 (1): 51–69.
- Carson, Thomas L. (2018) 'The Range of Reasonable Views about the Morality of Lying'. In Eliot Michaelson and Andreas Stokke (eds), *Lying: Language, Knowledge, Ethics, and Politics* (Oxford: Oxford University Press), 145–160.
- Day, J.P. (1993) 'Temptation', *American Philosophical Quarterly* 30 (2): 175–183.
- Dillof, Anthony M. (2004) 'Unravelling Unlawful Entrapment', *Journal of Criminal Law and Criminology* 94 (4): 827–896.
- Dworkin, Gerald (1985) 'The Serpent Beguiled me and I Did Eat: Entrapment and the Creation of Crime', *Law and Philosophy*, 4 (1), 17–39.
- Fletcher, George, P. (1978) *Rethinking Criminal Law*. (Boston: Little Brown).
- Foot, Philippa (1967) 'The Problem of Abortion and the Doctrine of Double Effect', *Oxford Review* 5: 5–15. Reprinted in Philippa Foot, *Virtues and Vices, and Other Essays in Moral Philosophy* (2nd ed., Oxford: Blackwell, 2002), 19–32.
- Ho, Hock Lai (2011) 'State Entrapment', *Legal Studies*, 31 (1), 71–95.
- Howard, Jeffrey W. (2016) 'Moral Subversion and Structural Entrapment', *Journal of Political Philosophy* 24 (1): 24–46.
- Hughes, Paul M. (2004) 'What is Wrong with Entrapment?', *Southern Journal of Philosophy* 42 (1): 45–60.
- Hughes, Paul M. (2006a) 'Temptation, Culpability and the Criminal Law', *Journal of Social Philosophy* 37 (2): 221–232.
- Hughes, Paul M. (2006b) 'Temptation and Culpability in the Law of Duress and Entrapment', *Criminal Law Quarterly* 51 (3): 342–359.

- Kant, Immanuel (1997) *Lectures on Ethics*, ed. Peter Heath and J. B. Schneewind, tr. Peter Heath. (Cambridge: Cambridge University Press).
- Mahon, James Edwin (2009) 'The Truth about Kant on Lies'. In Clancy Martin, ed., *The Philosophy of Deception* (Oxford: Oxford University Press 2009), 201–224.
- Marcus, Paul (2016) *The Entrapment Defense* (5th edition, LexisNexis).
- Miller, Seamus and John Blackler (2005) *Ethical Issues in Policing*. (Aldershot: Ashgate).
- Miller, Seamus, John Blackler and Andrew Alexandra (2006) *Police Ethics*. (2nd ed., Winchester: Waterside Press).
- Park, Roger (1976) 'The Entrapment Controversy', *Minnesota Law Review* 60: 163–274.
- Redmayne, Mike (2012) 'Exploring Entrapment'. In Lucia Zedner and Julian V. Roberts (eds), *Principles and Values in Criminal Law and Criminal Justice: Essays in Honour of Andrew Ashworth* (Oxford: Oxford University Press), 157–170.
- Stitt, B. Grant and Gene G. James (1984) 'Entrapment and the Entrapment Defense: Dilemmas for a Democratic Society', *Law and Philosophy* 3 (1): 111–132.
- Wells, Colin (2011) *Abuse of Process* (2nd edition, PLACE: Jordan Publishing).
- Wikipedia (2019) 'Q-ship'. Last modified 23 September 2019.
<https://en.wikipedia.org/wiki/Q-ship>.